

# Soybean gene express: plataforma para análise de expressão diferencial em bibliotecas subtrativas de cDNA

NASCIMENTO, L.S.<sup>1</sup>; SILLA, P.R.<sup>2</sup>; VITORINO, J.C.<sup>2</sup>; IWATA, M.<sup>2</sup>; MITANI, E. A.<sup>3</sup>; BINNECK, E.<sup>4</sup>

<sup>1</sup>Bolsista DTI, CNPq; <sup>2</sup>Universidade Estadual de Londrina – UEL; <sup>3</sup>Centro Universitário Filadélfia - UniFil. <sup>4</sup>Embrapa Soja, Caixa Postal, 231, 86001-970, Londrina, Paraná,  
e-mail: lukasnasc@cnpsa.embrapa.br

## Introdução

A utilização de técnicas de transcriptômica (ou genômica funcional) tem proporcionado o acesso cada vez mais amplo à funcionalidade do genoma da soja, permitindo a identificação de genes e rotas metabólicas relacionadas aos mais variados tipos de estresse. Várias técnicas são disponíveis para estudos de transcriptômica, entre elas, o sequenciamento de bibliotecas de transcritos completos (full length), de etiquetas de sequências expressas (ESTs - Expressed Sequence Tags), de SuperSAGE e de bibliotecas subtrativas.

Uma biblioteca subtrativa é um tipo de biblioteca de cDNA que representa clones existentes em um tecido ou organismo mas não em outro, obtidos em um determinado momento ou situação. Por exemplo, é possível construir uma biblioteca subtrativa que represente os genes diferencialmente expressos na comparação de um genótipo em condições normais de ambiente com o mesmo genótipo sob algum tipo de estresse (e.g., ferrugem), ou dois genótipos diferentes (e.g., resistente e suscetível) comparados sob as mesmas condições experimentais.

O seqüenciamento de uma biblioteca subtrativa gera um enorme volume de dados. Para a interpretação desses dados, faz-se necessária a implementação de ferramentas computacionais avançadas para filtragem, alinhamento, agrupamento e toda a gama de análises requerida para a extração da informação contida no universo dos dados produzidos. Isso, sem contar a necessidade de analisar conjunta e comparativamente os novos dados produzidos com os dados públicos disponíveis.

Este trabalho objetiva descrever o sistema computacional *Soybean Gene Express*, desenvolvido no Laboratório de Bioinformática da Embrapa Soja. O *Soybean Gene Express* foi desenvolvido com o objetivo de disponibilizar ferramentas de bioinformática para análise, armazenamento e visualização dos dados de bibliotecas subtrativas produzidas no projeto GenoSoja (Consórcio Nacional para Estudos do Genoma da Soja), possibilitando o rápido acesso às informações sobre a expressão diferencial e anotação automatizada dos genes modulados.

## Material e Métodos

O sistema *Soybean Gene Express* é composto por um banco de dados, *scripts* que automatizam o processamento dos dados e a alimentação do banco de dados, e uma interface Web para consulta e visualização das informações geradas pelo sistema.

O sistema é alimentado por um arquivo no formato Fasta, que contém os *reads* gerados pelo sequenciamento das bibliotecas subtrativas.

Para a construção do banco de dados, que tem a finalidade de armazenar e gerenciar os dados do *Soybean Gene Express*, foi utilizado o SGBD (Sistema de Gerenciamento de Banco de Dados) MySQL (MYSQL, 2010). Através da linguagem do MySQL, são controladas as informações de 5 tabelas no banco de dados do sistema. Abaixo está uma breve descrição de cada tabela que compõe o banco de dados:

- **Resultados\_blast:** tabela utilizada para armazenar informações referentes ao *conting/singlet, score, e-value, description, notes, GO* e BLAST.
- **Biblioteca:** tabela que contém as informações referentes às bibliotecas subtrativas submetidas no sistema.
- **Sub-biblioteca:** armazena as informações que descrevem os tratamentos experimentais relativos às bibliotecas.
- **ExpressaoDiferencial:** armazena os valores de quantidade de expressão dos genes em cada sub-biblioteca, bem como dados da expressão diferencial comparativa entre as sub-bibliotecas.
- **GeneOntology (GO):** armazena as informações referentes a anotação de ontologia dos genes. O GO fornece um conjunto de vocabulários estruturados para um domínio biológico específico que pode ser usado para descrever produtos de genes em um organismo.

A Figura 1 mostra um esquema dos passos executados para armazenar as informações no banco de dados.

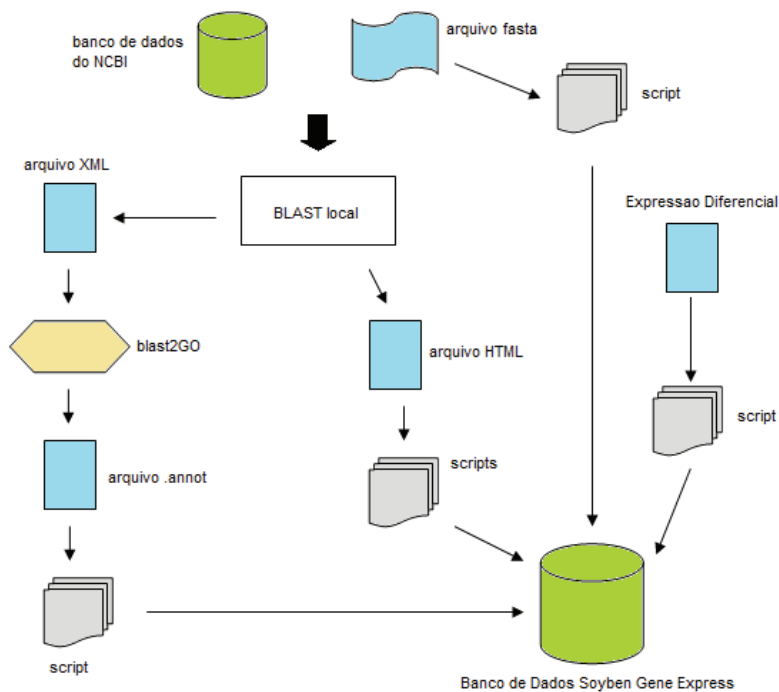


Fig. 1. Fluxo dos passos para o armazenamento das informações do sistema *Soybean Gene Express*.

Os *scripts* que fazem o processamento dos arquivos de entrada, análises das sequências e alimentação do banco de dados, foram desenvolvidos na linguagem de programação Python. Abaixo está a descrição dos *scripts*:

- **blast\_banco.py:** faz a leitura do arquivo gerado no BLAST e insere as informações necessárias no banco de dados.

- **cut\_blast\_nr\_v3.py:** divide o arquivo gerado no BLAST em vários sub-arquivos referentes aos contigs e singlets.
- **sequencia\_blast.py:** faz a leitura da sequência de cada contig/singlet do arquivo FASTA que foi utilizado pra rodar o BLAST e o insere no banco de dados.
- **subbiblioteca.py:** faz a leitura do arquivo e armazena as sub bibliotecas no banco de dados.
- **expressaodiferencial.py:** faz a leitura do arquivo e armazena, no banco de dados, a quantidade de *reads* encontrados referentes ao contig/singlet e a sub biblioteca.
- **go.py:** faz a leitura do arquivo de saída gerado pelo blast2GO e armazena estas informações no banco de dados.

Além dos *scripts* desenvolvidos, vários *plug-ins* Javascript foram utilizados para validação de dados de formulário e formatação da tabela, como exemplo o *jQuery* (JQUERY, 2010) e *dataTable* (DATATABLES, 2010).

Para o desenvolvimento da interface de consulta e visualização dos dados (Fig. 2 e 3), foi utilizada a linguagem de programação para web PHP (PHP, 2010).

This page is intended to consult the results obtained from the Blast.

Select the library:  (1)

Select search field:  (2)

Keyword:  (3)

(4)

**Blast Results**

Total de registro(s): 3 (a)  
 Show  entries (b)

Contig/Singlet	Score	E-value	Description	Notes	GO	Blast
BD4_414-26_372123	122	2e-26	gb ADD6320L.1 cytochrome b6-f complex subunit IV [Phoenix ...	Notes	GO:000775 - cytochrome b6 f complex subunit iv GO:0016491 - cytochrome b6 f complex subunit iv GO:0045156 - cytochrome b6 f complex subunit iv GO:0045158 - cytochrome b6 f complex subunit iv GO:0042951 - cytochrome b6 f complex subunit iv GO:0016021 - cytochrome b6 f complex subunit iv GO:0006507 - cytochrome b6 f complex subunit iv GO:0022904 - cytochrome b6 f complex subunit iv	Blast
Contig00472	279	9e-74	gb ADD31082.1 cytochrome b6 protein [Gunnera manicata]	Notes	GO:0016491 - cytochrome b6 GO:0005506 - cytochrome b6 GO:0006110 - cytochrome b6 GO:0045158 - cytochrome b6 GO:0016021 - cytochrome b6 GO:0022904 - cytochrome b6 GO:0015679 - cytochrome b6 GO:0006535 - cytochrome b6	Blast
Contig0053L	119	1e-25	ref YP_538795.1 cytochrome b6 [Glycine max] >gi 122202460 ...	Notes	GO:0016491 - pe tb GO:0005506 - pe tb GO:0006110 - pe tb GO:0045158 - pe tb GO:0016021 - pe tb GO:0022904 - pe tb GO:0015679 - pe tb GO:0006535 - pe tb	Blast

(c) (d) (e) (f) (g) (h) (i)

**Fig. 2.** Resultado de anotação BLAST e GO com as principais informações - (1) Selecione o tipo de biblioteca, (2) selecione o campo de consulta, (3) digite a palavra chave para busca, (4) botão buscar, (a) total de registros encontrados, (b) escolha quantidade de linhas para mostrar na tela, (c) nome do *contig* ou *singlet* e *link* para o relatório da sequência, (d) score, (e) e-valor, (f) descrição do *hit* com *link* para o site do NCBI (NCBI, 2010), (g) anotação manual, (h) gene *ontology* (BLAST2GO, 2010) e (i) link para relatórios BLAST.

The screenshot displays the 'Soybean Gene Express' web interface. At the top, there are logos for 'Soybean Gene Express', 'NTBIOS', and 'Embrapa Soja'. Below the logos is a navigation bar with links: HOME | BLAST RESULTS | DIFFERENTIAL EXPRESSION | CONTACT | HELP | BLAST2GO.

The main content area is titled 'Differential Expression'. It includes a search form with three fields: 'Select the library:' (containing 'Ferrugem'), 'Select the sub-library:' (containing 'BRS231Bulk1\_12\_24\_48'), and a 'Search' button. Below the search form, there are labels (1), (2), and (3) pointing to the respective fields.

Below the search form, there is a section titled 'Differential Expression' with a 'Show the sequences selected' button (a). A legend (b) shows 'Vale mais express' (green) and 'Second vale mais express' (blue). A 'Show 10 entries' dropdown (c) is present. A search bar (d) is also visible.

The main part of the interface is a table (e) showing differential expression results. The table has columns for 'Fold Change', '(Contigs/Singlets)', and several sub-library identifiers: BRS231Bulk1\_12\_24\_48, P1561356\_024, Embrapa48\_024, BRS231Bulk2\_72\_96, Embrapa48\_192, BRS231Bulk3\_196, and P1561356\_192. The table lists 13 rows of data, with the first row having a 'Fold Change' of 6 and 16 contigs in the BRS231Bulk1\_12\_24\_48 library. Below the table, there is a 'Showing 1 to 10 of 413 entries' label and a pagination control (f) with buttons for 'First', 'Previous', '1', '2', '3', '4', '5', 'Next', and 'Last'.

**Fig. 3.** Expressão Diferencial - (1) escolha da biblioteca para consulta, (2) escolha da sub-biblioteca, (3) botão buscar, (a) lista as sequência dos itens selecionados na tabela, (b) legenda dos valores mais expressos, (c) escolha da quantidade de linhas da tabela para mostrar na tela, (d) busca informações da tabela, (e) lista a quantidade de registros encontrados e (f) possibilita navegar entre as páginas da tabela.

## Resultados e Discussão

O *Soybean Gene Express* pode ser acessado pelo endereço <http://bioinfo.cnpso.embrapa.br/soybeangeneexpress/index.php>. O sistema possui uma interface de consulta amigável, de fácil navegação de forma a facilitar a pesquisa do usuário. O usuário poderá pesquisar os resultados de anotação BLAST e GO, os resultados de Expressões Diferenciais, bem como fazer anotações manuais para cada gene de forma intuitiva e objetiva. Nas Figuras 2 e 3 são mostrados exemplos de buscas que podem ser realizadas no sistema.

## Conclusões

O *Soybean Gene Express* compreende um sistema que agrupa em uma única ferramenta as informações necessárias para acessar a modulação da expressão, bem como interpretar o significado dos genes modulados em experimentos que envolvem a utilização de bibliotecas subtrativas. Isso facilita sobremaneira o trabalho do pesquisador, estudante ou técnico de laboratório na identificação dos genes cuja expressão é influenciada pelos tratamentos estudados.

Este sistema mostra os benefícios de se utilizar ferramentas computacionais para a disponibilização de grande quantidade de informações de maneira a agilizar a pesquisa e a interpretação de dados de estudos genômicos.

## Referências

- BLAST2GO. Disponível em: <<http://www.blast2go.org/>> Acesso em: mar. 2010.
- DATATABLES. Disponível em: <<http://www.datatables.net>> . Acesso em: fev. 2010.
- JQUERY. Disponível em: <<http://www.jquery.com>> . Acesso em: fev. 2010.
- MYSQL. Disponível em: <<http://www.mysql.com>> . Acesso em: jan. 2010.
- NBCI. Disponível em: < <http://blast.ncbi.nlm.nih.gov>> . Acesso em: fev. 2010.
- PHP. Disponível em: <<http://www.php.net>> . Acesso em: jan. 2010.